DOCUMENT RESUME

ED 442 813 TM 031 233

AUTHOR Pommerich, Mary; Burden, Timothy
TITLE From Simulation to Application: Examinees React to Computerized Testing.
PUB DATE 2000-04-00
NOTE 37p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 25-27, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adaptive Testing; *Attitudes; Comparative Analysis; *Computer Assisted Testing; *High School Students; High Schools; Problem Solving; Simulation; *Test Format; Test Items; *Testing Problems
IDENTIFIERS Paper and Pencil Tests

ABSTRACT

A small-scale study was conducted to compare test-taking strategies, problem-solving strategies, and general impressions about the test across computer and paper-and-pencil administration modes. Thirty-six examinees (high school students) participated in the study. Each examinee took a test in one of the content areas of English, Mathematics, Reading, and Science. In spite of the small sample, observations from the study highlight issues test developers might want to consider in determining how to present a test. Several factors were identified that might lead an examinee to respond to more than just item content when giving an answer: page and line breaks, passage and item layout features, highlighting, and item characteristics. Other factors include navigational features such as scrolling, item review, item preview, and omit capability. Examinee characteristics contributed to many of the observed mode effects, especially examinee carelessness. Care should be taken to ensure that the examinee is responding to item content only and not to inherent features associated with the test administration mode. (Contains 16 references and 16 tables.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 442 813

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

M. Pommerich

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

From Simulation to Application: Examinees React to Computerized Testing

Mary Pommerich
Timothy Burden

ACT, Inc.

TM031233

BEST COPY AVAILABLE

Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA, April 2000.

Acknowledgements

The authors gratefully thank David Duer, Beth Gehring, Ann Gordon, Beth Hill, Toni McCarthy, Valerie Nelson, Jim Patterson, and Phillip Zell for their countless contributions toward defining hypotheses, developing interview questions and protocols, and conducting the interviews.

From Simulation to Application: Examinees React to Computerized Testing

The advent of computerized testing introduces the issue of how to present test items in a medium that differs substantially from the conventional test booklet used in paper and pencil testing. Years of research with paper and pencil tests have led to decisions about how to format and present passages and items within a test booklet. Formatting practices applied to booklet presentation might or might not be appropriate for a computer presentation of the same material. Booklet formatting decisions give us a starting point for formatting decisions for computerized presentation. But it is not clear whether the expectations of examinee performance and behavior based on research with conventional paper and pencil tests will apply to the less understood setting of computer administered tests.

Ideally, mode of administration, whether paper and pencil or computer, should not be a factor in how an examinee responds to a test item. Responses to an item should be dictated by item content only, and examinees both within modes and across modes should react to the item content rather than the features inherent to presenting the item in that mode. Due to differences in the administration media, it might not be feasible (or possible) to present the same form of a test in exactly the same manner in a test booklet and on a computer screen. For discrete item tests such as a mathematics test, an examinee might see multiple questions across a two-page spread in a booklet presentation. But in a computer presentation of the same material, it might be best to present only one item at a time on the computer screen. For passage-based tests such as a reading test, an examinee might see a passage in its entirety and some number of related questions across the two-page spread within their view. But in a computer presentation of the same material, it might be best (or possible) to present only a portion of the passage and questions at a time on the computer screen.

Examinees might have innate reactions to how a test is structured in the presentation media. If an examinee takes a passage-based test presented in a booklet, they might be more inclined to read the entire passage first before looking at the questions. Whereas, an examinee taking the same test presented via computer, might be more inclined to start answering the questions directly without first reading the entire passage. Individual test-taking styles dictate to some degree how examinees will approach a test. Because of subtle differences in test presentation across administration modes, care must be taken to ensure that examinees respond to item content only. Examinee item responses should not be affected by features that are an

artifact of the method of presentation. This is true regardless of whether a testing program employs only one mode of testing or employs both computerized and conventional testing.

Take the case of an item that references information occurring before a page break in a booklet. Some individual examinees might read information on the previous page, whereas others might not because of the page break. Thus, an examinee's response might be inherently affected not just by the item content, but also by the method in which the test and items are presented. With a computer presentation of the same material that requires examinees to scroll through the passage, the page break factor is removed from that item (although new presentation factors might arise). For a testing program that employs testing in computer and paper and pencil modes, such presentation differences could contribute to mode effects. Depending upon the administration mode the examinee chooses, the examinee might or might not be affected by the page break. There is no page break issue for the computer, and depending upon the examinee's individual characteristics, there might or might not be a page break issue for the paper and pencil administration. Thus, there is a potential source of difference in performance across the two modes of administration.

Computer-based versus computer adaptive tests add other presentation factors to the mix, namely, the ability or inability to review, preview, and omit items. Some computer-based tests are essentially a computerized presentation of a paper and pencil test, and could allow the same freedom of movement as the paper and pencil test. In a computer adaptive test, where items are selected for administration based on the examinee's performance up to that point, allowing review, preview, and omits is a difficult (or impossible) task. The practice of many computer adaptive tests has been to not allow the freedom of movement that is inherent to paper and pencil testing. This inhibition could also contribute to mode differences across paper and pencil and computer administered tests. Computer adaptive tests also increase the possibility that responses to an item might be influenced by the content of other items, by the position in which the item is presented, or by previous exposure to the item.

A number of simulation-based research studies have been conducted at ACT as part of the process of developing test administration procedures for computerized tests (Davey & Nering, 1998; Davey, Nering & Thompson, 1997; Fan, Thompson, & Davey, 1999; Hsu, Thompson & Chen, 1998; Nering, Miller & Davey, 1999; Thompson & Davey, 1997; Parshall, Davey & Nering, 1998; Reckase, Thompson & Nering, 1997; Thompson & Davey, 1999;

Thompson, Davey & Nering, 1998; Thompson, Nering & Davey, 1997). Although these simulations evaluate the technical aspects of procedures for administering a test on computer, they cannot predict how examinees will react and function during that administration. Ensuring psychometric quality of computerized tests does not always ensure psychometric success, because we do not know how examinees will react to the test administration procedures, both psychometrically and psychologically.

A number of real-data-based research studies have been conducted to examine score comparability across computer and conventional administrations of the same items. Spray, Ackerman, Reckase, and Carlson (1989) compared total test score across computerized and paper and pencil presentation of tests for the Marine Corps. The computerized tests were administered to allow the same freedom of movement as the paper and pencil tests. They did not find mode differences, and attributed those findings to the freedom of movement they allowed, and that the test items appeared on screen exactly as they appeared in the paper and pencil version. The test items contained minimal text and no figures, graphics, or schematics were used. Mazzeo, Druesne, Raffeld, Checketts, and Muhlstein (1991) found mode effects (differences in average scores) across computer and paper and pencil presentations of two CLEP tests. Based on their findings and comments from study participants, they made modifications to the computer presentations that eliminated the mode effects in one of the tests, but not the other. Examinees were also questioned as to their computer familiarity. Schaeffer, Reese, Steffen, McKinley, & Mills (1993) found no substantive item-level mode differences in paper and pencil and computer presentations on the GRE. Examinees were questioned as to their reactions to the computer-based test on issues such as computer experience, using the interface tools, scrolling, item review, and omitting practices. Parshall and Komrey (1993) studied the effect of examinee demographics, computer experience, and review and omit strategies on total score across modes, but the relationships between examinee characteristics and mode effects were weak. Neuman and Baydoun (1998) discussed several potential sources of mode differences (different stimulus presentation or response procedures, requirement of different motor skills, or computer anxiety) but did not examine any sources in their evaluation of the equivalence of a speeded clerical test battery.

The focus of most real-data-based research has been on determining whether or not scores are comparable at a total score level. Further, any feedback from examinees participating

in the previous studies appeared to be solicited for the test overall, rather than at the item level. Where performance differences were found at an item level, researchers did not generally look in depth at explanations of the source of mode differences. Merely identifying that items performed similarly or differently across modes did not offer an account for why they performed similarly or differently, or whether there were presentation features within a mode that caused examinees to react to more than just the item content. Given our principle that responses should be dictated only by item content, we are interested in identifying and minimizing presentation differences that contribute to mode differences at the item level.

A previous study at ACT administered the same items either conventionally or on computer to randomly equivalent groups of examinees. Results suggested that while mode effects were slight at the total test level, certain items showed larger performance differences in one direction or the other. But determining the causes of those differences based purely on the test data at hand proved a difficult task. Hypotheses about the causes were formulated, but not confirmed. To account for differences, we performed a qualitative study that enabled us to study in depth what examinees did when taking a computerized or paper and pencil version of a test, what their approach to taking the test was, and how that approach was influenced by the presentation of the test and items. This study focused at the item level, and attempted to identify presentation features that might cause mode differences.

Method

Study Design

A small-scale study was conducted to compare test-taking strategies, problem-solving strategies, and general impressions about the test across computer and paper and pencil administration modes. Thirty-six examinees participated in the study. N-counts by test and mode are given in Table 1. Each examinee took a 20-minute test in one of the content areas of English, Mathematics, Reading, and Science. The same items were administered under each administration mode, the only differences being those necessary to present items on the computer rather than in a printed booklet. Multiple content areas were included to accommodate presentation features unique to each test. Following each test administration, examinees were interviewed extensively as to their approach to solving a subset of the test questions.

Table 1. Number of Examinees by Test Mode and Content Area

Content Area	Computer	Conventional
English	5	3
Mathematics	5	5
Reading	5	4
Science	4	5
Total	19	17

The item subset and interview questions were selected based on *a priori* hypotheses about how presentation features might lead examinees to perform differently across administration modes, or to use different problem-solving strategies and approaches across administration modes. For the selected test items, examinees were asked to recreate what they did to get the answer they gave, and then were asked comprehensive questions addressing the examinee's process of navigating through the test. The questions were developed to address the presentation features of interest and were designed to determine how the examinee interacted with and reacted to the presentation features in answering the item.

All attempts were made to ask questions that would allow evaluations of how the presentation features affected performance in each administration mode, without leading examinees to give answers that corresponded to our hypotheses or to answer in a way we anticipated they would answer. Questions about specific items were then followed by more general questions about test-taking strategies, opinions about the structure of the test, and reactions to the test interface. The interview finished with questions about the examinee's academic experience, previous test-taking experience, and computer experience. All questions were written to be as non-directional in scope as possible, so that examinees would freely choose a direction as opposed to being subtly guided to choose a direction by the wording of the question.

A follow-up interview was used so that examinees could take the test uninterrupted, under timed conditions similar to the usual administration of the test. The test length was kept to 20 minutes, with the hope that examinees would be able to remember the test questions and what they had done to answer the questions. Test lengths were 30 items for English, 15 items for Mathematics, 20 items for Reading, and 19 items for Science. Both computer and paper and pencil examinees were prompted when there were 5 minutes of testing time remaining. Computer examinees took a short tutorial that demonstrated the functions necessary to take the

computerized test. The entire testing and follow-up interviews were conducted in 2-hour sessions for each examinee. Examinees taking the computer administration were videotaped both taking the test, and throughout the interview. Examinees on the paper administration were not videotaped, although a video camera was present, and they were told they were being taped. Audiotapes were made of both the computer and paper administrations, from which transcripts of the sessions were created.

Because of the length of time per session and the restriction of one-on-one interactions with the examinee, the sample sizes were limited to a handful of examinees per administration mode per content area. As a result of the restricted sample sizes and the non-quantitative nature of the information collected, we attempt to present the findings as more observations than conclusions. Any conclusions we do draw are truly speculative in nature, and are specific to the group of examinees studied. We do believe that if one examinee demonstrates a certain behavior, it is possible that others might also exhibit the same behavior, although we cannot predict the extent to which that behavior might occur.

Examining in depth examinee responses to the interview questions provided us with some ideas as to how an individual examinee might interact with the presentation features of each administration mode both knowingly and unknowingly, and how an individual might react psychologically to the features inherent to the presentation mode. Reactions might occur at two levels. One is the examinee's own recognized reaction to the testing situation, namely, how the examinee feels about the test. The second type of reaction might be a subtler interaction with the manner of presentation of the test. In many cases, the examinee might not make a conscious choice about how they interact with the interface, but rather might react innately to the manner in which the test is presented.

Sample Solicitation and Description

Students were recruited for study participation by advertising in the local newspaper and e-mail solicitation of ACT staff members. The study was conducted in August, 1999. Rising juniors and seniors for the 1999-2000 school year were solicited. Examinees were paid a stipend for participating in the study. Parental consent was required to participate in the study. Consent from the examinee was also obtained the day of testing. Students signed up for a testing time on a first-come, first-serve basis. There was no random assignment of students to test content or mode, but rather, interviewers were assigned according to scheduling convenience. Four

interviewers participated, one each for the English, Reading, Mathematics, and Science tests. The interviewer for a content area conducted the sessions for both computer and conventional administrations. All possible steps were taken to ensure that participants in the study did not know the person conducting their test and interview session.

The sample consisted of 17 males and 19 females. The participants included 28 Caucasian-Americans, 2 African-Americans, 2 Asian-Americans, and 1 multi-racial examinee. Three examinees chose not to give their ethnicity. Fifteen of the examinees were rising seniors, 19 were rising juniors, and 2 were rising sophomores[1]. Thirty examinees attended high schools within larger school districts. Six of the examinees attended high schools in smaller, more rural areas. The average reported grade-point-average was 3.44 for computer administration, and 3.49 for paper administration. Examinees demonstrated various levels of computer experience. Thirty-four examinees reported having a computer in their home. Of those 34, 14 reported using it daily, 12 reported using it often, six reported using it infrequently, while one reported no usage. There was one unknown usage. Two examinees reported having no computer in the home. However, both of those examinees reported using computers at school. One characterized his computer skills and knowledge as about the same as other kids his age, whereas the other examinee characterized her skills as less than other kids her age. The latter examinee, however, expressed an interest in taking tests on computer rather than by paper and pencil administration.

Description of Test Presentations

English

The English test consisted of two passages containing underlined words and phrases, with 15 multiple-choice items in each passage. For most items, examinees were instructed to choose the response option that best expressed the idea, made the statement appropriate for standard written English, or was worded most consistently with the style and tone of the passage as a whole. These types of items had no stimulus associated with them (i.e., there were only response options, and no preceding question). For some items, there was a stimulus present that asked a question about the underlined portion in the passage. Examinees were instructed to choose the best answer to the question.

---

[1] One rising sophomore took the English test, while the other took the Reading test. Both of those tests were deemed suitable for a student that might not have had the recommended coursework prior to testing.

In the booklet presentation, the passage and items were presented jointly on a page. The passage was presented in the left half of the page, while the items were presented in the right half of the page. The passages and accompanying items occupied about two booklet pages each. Examinees were able to move freely throughout all English passages and items in the booklet while taking the English test. They could respond to items and passages in any order, and were not required to give responses to all items. Similar rules of movement between items and passages held for the Mathematics, Reading, and Science paper and pencil tests. Within a single test, examinees were allowed to move freely throughout the test.

In the computer presentation, the passage and items were presented jointly on the screen, with the passage on the left half and the items on the right half of the screen. The complete underlined portion for each item was highlighted in the passage window. The examinee had to scroll through the passage to see the passage in its entirety. Items were presented one at a time, and the examinee had to select each item to respond to it, with the exception of the first item, which showed up on screen at the start of the passage. The passage automatically scrolled for examinees when they selected an item that was not visible in the passage window. Within a passage, examinees were allowed to answer items in any order. They were required to answer all items prior to moving on to the next passage. Once an examinee completed a passage and moved on to the next passage, they were not allowed to return to the previous passage. Also, passages were presented one at a time, so that examinees could not see the next passage until they proceeded to it. A similar presentation of the passage and item windows was used with the computerized Reading and Science tests, along with the same rules for moving between items and passages.

Mathematics

The Mathematics test consisted of 15 discrete multiple choice items. Some items contained figures. Examinees were allowed to use a calculator on the test. In the booklet presentation, the items appeared sequentially in the booklet. Examinees were allowed to write in the test booklets to solve the problems. In the computer presentation, the items were presented one at a time. Examinees were required to give a response to the item before moving on to the next item. Examinees could only see the current item on-screen, and were not allowed to go back to previous items, or see the next item until they proceeded to it. Computer examinees were provided with scratch paper and pencils to solve the problems.

Reading

The Reading test consisted of two passages with 10 multiple-choice items on each passage. Examinees were instructed to read the passage and choose the best answer to each question. In the booklet presentation, the reading passage was presented first in its entirety, in two columns per page. The passages were followed by the test items. The passages and accompanying items occupied about two booklet pages each. The computer presentation for Reading corresponded to that described for the English test. Items on the Reading test generally fell into two types: questions that required a global understanding of the passage and questions that required knowledge of specific information given in the passage. For global questions, examinees typically had to make an inference from what they had read to answer the question. Some of the items had line references associated with them (i.e., the item stimulus contained the number of a line or lines in the passage to which they were directed to read).

Science

The Science test consisted of three passages with varying numbers of multiple-choice items per passage (5-7 items). Some passages contained figures and tables. In the booklet presentation, the passage was presented first in its entirety, in two columns per page. The passages and accompanying items occupied about two booklet pages each. The passages were followed by the test items. The computer presentation for Science corresponded to that described for the English test, with the additional feature that some figures and tables within the passage were enlargeable and moveable.

Hypotheses and Findings

How an examinee approaches an item and reacts to the presentation of an item and test has to do with that individual examinee, in terms what their usual test-taking practices and strategies are, and how those tendencies interact with the presentation characteristics for the item. On the computer side, an examinee's reaction to the item and test presentation might also be related to their computer experience and their level of fluency with the computer. In this section, we will discuss findings and suppositions pertaining to item-specific issues on each test. For each of the tests, we will identify item characteristics and related passage characteristics (for passage-based tests), and present hypotheses about several possible sources of differences in performance across administration mode.

For the handful of examinees taking each administration mode within each content area, we will present examples of how they interacted with the testing interface. We will also discuss more global issues that were relevant across all four tests. We reiterate that we do not intend to draw any strong conclusions about what other examinees might do in the same circumstances. But if one examinee exhibits a behavior, other examinees might also exhibit that behavior, so that we should be aware of that potential behavior in making formatting and presentation decisions for each administration mode. We understand that on the basis of such small sample sizes, it is impossible to predict what will happen with large numbers of examinees. It does give us an idea, however, of what we might expect some examinees to do and how we might expect them to react.

Item-Specific Issues

English

Based on previous experience, we anticipated that the computerized administration of the English test might favor computer examinees overall, so that computer examinees might perform better on average on the test than the paper and pencil examinees. Our hypothesis was that highlighting might have given computer examinees greater focus on the underlined portions, so that they read them in their entirety and were able to better associate them with the corresponding stimulus and item. At an individual item level we anticipated that some items might favor computer examinees, while others might favor paper and pencil examinees. We hypothesized that there were differences in presentation of the passages and items across modes that might contribute to those differences (issues such as page breaks and page layout). We also hypothesized that those differences might interact with examinee test-taking characteristics (such as whether the examinee read the item stimulus or not, and the order in which the examinee read response options). Careless examinees might not read the test instructions carefully, and thus might be unaware that they might need to read the stimulus on items where it is present. Careless examinees might also be more inclined to read the options only up until the point where they pick their answer. More careful examinees might purposely choose the same strategy as a timesaving device. Sometimes we identified multiple factors within an item, which we anticipated might counteract with each other and result in no difference in performance across modes.

We asked examinees about their experience in answering 11 of the 30 English items. The examinees' interactions with the computer and booklet interfaces will be discussed for issues of page breaks, passage layout, items with a stimulus, and highlighting. Table 2 presents a summary of examinee performance across items on the English test, by examinee ID. A 'C' in the ID represents a computer examinee, while a 'P' represents a paper and pencil examinee. Examinees will be referred to by these IDs on occasion and referenced with regard to their reaction to test and item-specific issues. The shaded items are items that will be discussed relative to a test-specific issue.

Table 2. English Item Responses and Key

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
EC1	C	C	A	C	A	C	C	B	C	D	B	D	D	A	A
EC2	B	C	B	B	A	C	C	A	D	D	B	C	D	A	A
EC3	C	C	B	A	A	C	C	C	D	D	C	A	A	A	A
EC4	B	C	B	D	A	C	C	C	B	D	B	C	D	A	A
EC5	C	B	C	D	A	C	C	A	B	D	B	C	D	A	D
EP1	D	C	B	C	A	C	C	A	B	D	A	D	D	A	B
EP2	C	B	B	C	A	B	C	A	B	D	B	C	C	A	D
EP3	B	C	B	D	A	B	C	A	D	D	B	C	D	A	A
Key	B	C	B	D	A	C	C	A	B	D	B	C	D	A	A

ID	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Total
EC1	D	B	D	A	B	A	D	A								13
EC2	B	A	D	B	B	A	B	C	B	B	D	D	D	A	D	27
EC3	A	C	C	A	B	C	B	C	B	B	C	D	B	D	D	16
EC4	B	A	D	A	B	A	B	C	B	B	D	D	D	A	D	29
EC5	B	B	A	D	B	D	B	A	A	D	D	D	D	D	D	18
EP1	A	C	D	A	B	A	B	A	A	B	C	D	A	B	C	17
EP2	B	C	B	A	B	A	B	C	B	B	D	D	D	A	D	22
EP3	B	C	B	A	B	A	B	C	B	B	D	D	D	A	D	26
Key	B	A	D	A	B	A	B	C	B	B	D	D	D	A	D	30

Page Breaks. Item 6 presented an example of a page break in the booklet presentation versus no page break in the computer presentation. We anticipated the item might favor computer examinees, if they were more inclined to read the entire sentence containing the underlined portion. In the booklet, a page break occurred in the middle of the sentence containing the underlined portion, so that the beginning of the sentence was contained on one page, while the rest of the sentence was completed on another. The underlined portion consisted of two words, and was contained on the second page of the passage. In the computer presentation, the underlined portion was typically located in the middle of the passage window

BEST COPY AVAILABLE

(depending upon how the examinee maneuvered throughout the interface, the location of the underlined portion might have differed). Table 3 summarizes the results for this item.

Table 3. Results for Item with Page Break Issue

	Computer	Paper
Read entire sentence	4	2
Answered item correctly	5	1
Number Taking Item	5	3

All five computer examinees answered the item correctly. Except for one examinee, the computer examinees all indicated that they read the entire sentence. Examinee EC4 indicated he did not read the entire sentence, but that he had read enough of the sentence to know what he had to answer. Two of the three paper examinees indicated they read the entire sentence; one did not read the entire sentence. Even for paper and pencil examinees that read the entire sentence, the page break might have been distracting to their performance on the item. Examinee EP3 (who did read the entire sentence and answered the item incorrectly) expressed a strong dislike of questions involving a page break because he thought that having to read across pages made questions harder. Although these results do not give us any certainty, there might be an inclination for paper and pencil examinees not to read information on the preceding page, and page breaks might be a distraction for paper and pencil examinees that do choose to read the preceding information.

Item Stimulus. Item 10 was the first item to contain a stimulus that the examinee was supposed to read for instructions on how to respond. We anticipated this item might favor paper and pencil examinees if they were more inclined to read the stimulus. In the booklet presentation, the underlined portions were always lined up with top of the item (whether there was a stimulus or not). This sometimes required the use of white space, or gaps, between adjacent underlined portions of the passage. In the computer presentation, the item position was fixed in the item window, and the underlined portion in the passage window was not aligned with the top of the item in the item window. Also, items were numbered at the top of the item window, rather than right next to the item within the item window. Hence, there was some concern that examinees on the computer side might be less inclined to read the stimulus than paper and pencil examinees. Table 4 shows the summary of results for Item 10, and for Items 18 and 22, which also contained a stimulus.

BEST COPY AVAILABLE

On all items, the examinees displayed varying reactions to the stimulus. On Item 10, all answered correctly, but not all read the stimulus for this item. It might have been possible for some examinees to infer the right answer without reading the stimulus. For example, Examinee EC5 did not read the stimulus, but indicated she was able to make an inference from the options and the content of the passage up to that point. Less astute examinees might not have been unable to answer correctly without reading the stimulus. Examinee EC2 indicated he did not read the stimulus initially, but upon reading the stimulus, he changed his approach and answered according to the stimulus rather than the non-stimulus format. Examinee EC3 read the stimulus only after reading the response options. Examinee EP3 explicitly stated his recognition of the stimulus as a different type of instruction for the item. Examinee EP1 seemed confused by the nature of the question, and indicated that she guessed.

Table 4. Results for Items with Stimulus Issues

	Item 10		Item 18		Item 22	
	Computer	Paper	Computer	Paper	Computer	Paper
Read stimulus initially	3	2	2	2	1	2
Read stimulus but not initially	1	0	2	0	2	0
Did not read stimulus / Did not read stimulus carefully	1	1[?]	1[?]	1	2[??]	1
Answered item correctly	5	3	3	1	4	3
Number Taking Item	5	3	5	3	5	3

[?] It is not clear from examinee comments whether the examinee read the stimulus or not.
[??] One examinee indicted he did not read the stimulus carefully, the other might or might not have read the stimulus.

On Items 18 and 22, the examinees displayed varying strategies for dealing with stimulus-type items. The examinee recreations of their strategies for answering these items sometimes revealed what appeared to be timesaving approaches that affected whether they read the stimulus or not. Examinee EC5 showed some indication that she read the stimulus for given items only if it was not obvious from the response options what she was supposed to do. Examinee EC3 admitted to reading the underlined portion of Item 18 first, followed by the response options, only then followed by the stimulus, as a strategy for moving through the test faster. His intention was so that "when I read the [stimulus] I would have already an idea of what type of answer I wanted." This might lead to wrong answers if the option the examinee chooses does not correspond to what the stimulus is asking for. Examinee EC2 indicated he did not read the stimulus for Item 18 at first, but "then I realized it [the item] wasn't grammatically

incorrect without reading the [stimulus] first. So then I read the [stimulus]." Examinee EC1 indicated that he did not read the stimulus initially for Item 22 because he did not think that it was a stimulus-type question. This suggests that some examinees might look at response options first, and use the option format to determine whether or not it is necessary to look for and read a stimulus. Comments from the three paper and pencil examinees suggest that none of them read the stimulus carefully for Item 18. Examinee EP2 indicated she did not read the stimulus for Item 22 either, and that she answered essentially by instinct, stating: "I'm learning I need to look at the [stimulus]."

The results of Table 4 do not necessarily suggest that the computer examinees performed worse than the paper and pencil examinees on these items. Examinees across both modes indicated that they did not read the stimulus or did not read it carefully. It does show, however, that examinees take very different approaches to stimulus-type items. So that if there is a greater incidence of examinees not reading the stimulus on the computer presentation—because of the layout of the passage and item windows—than not reading the stimulus on the booklet presentation, this is a likely source of mode differences.

Highlighting. For several items, both computer and paper and pencil examinees were asked whether they thought their performance would have been different if the underlined portion was (paper) or was not (computer) highlighted. The computer examinees in general thought that highlighting helped their performance on the test, whereas the paper and pencil examinees generally thought that highlighting would not have made a difference in their performance. On Item 13, three of the five computer examinees thought that highlighting helped their performance. The remaining examinees (2 computer, 3 paper and pencil) thought that highlighting did not or would not have made a difference in their performance. Examinee EC1 thought that he might have missed a word in the underlined portion without highlighting. Examinee EC3 thought that highlighting helped give focus: "If it's not highlighted…you have a bit more trouble focusing on that one part." Examinee EC4 indicated that "it made it a little bit easier to see…what I was working on rather than looking through for the number." On Item 14, similar sorts of thoughts were expressed. Examinee EC1 thought it might take longer to answer the question without highlighting. Examinee EC4 thought that highlighting helped a little in terms of "…not taking so long to narrow down where it is [the underlined portion]…after

looking through the [passage]." Examinee EP2 thought that had the underlined portion been highlighted in the booklet, it might have stood out more.

Item 17 also demonstrated a potential "focus" effect for computer examinees due to highlighting. The underlined portion in Item 17 used the term "dualities", a word that many examinees might not have been familiar with. The response options all included the word, and variations on how to state the underlined portion. The correct response was "No change". We anticipated that highlighting on the computer mode might make examinees more likely to focus on the underlined portion as a viable option, because they did not understand the difficult word, and what the sentence was saying. Two of the five computer examinees answered the item correctly. None of the three paper and pencil examinees answered the item correctly. It is unclear what role, if any, highlighting played in the examinees' performance on this item. EP2 did indicate, however, that "maybe it [highlighting] would have made me think 'Oh, maybe I should focus on what they put first'... Sometimes I think I want to change things too much. Like I want to go and see what the changes are before I just look and see...maybe...what they have is probably maybe right."

Mathematics

Based on prior experience, we anticipated that Mathematics might show few differences at the item level across presentation mode. Because it was a fairly straightforward task to match the computer presentation of the study items to the booklet presentation, we anticipated that any favoritism, where existing, might be slight and not occur in a consistent direction over all items. We anticipated, however, that paper and pencil examinees might show more work than computer examinees, because of the greater ease of writing in a test booklet than in switching between the mouse and pencil to write on scratch paper. For problems with figures, we anticipated that paper and pencil examinees might mark on figures at a greater rate than computer examinees (as computer examinees would have to draw the figure on their scratch paper first). On problems with no figure, but where a figure might be helpful in solving the problem, we anticipated that computer examinees might be more likely to draw a figure than paper and pencil examinees, because they would be in the habit of drawing figures for previous items. We anticipated that paper and pencil examinees might not consider drawing a figure, since figures are typically drawn for them in the booklet.

We asked examinees about their experience in answering each of the 15 Mathematics items. We focus here solely on issues related to using scratch paper on the computer versus writing in the test booklet to solve problems. Table 5 presents a summary of examinee performance across items on the Mathematics test, by examinee ID. The total scores indicate that most of the examinees performed poorly on the Mathematics test.

Table 5. Mathematics Item Responses and Key

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
MC1	D	B	D	E	B	B	A	A	A	A	C	A	A	E	C	2
MC2	E	D	D	A	B	C	C	B	C	B	E	A	B	A	C	11
MC3	E	D	D	A	D	D	E	B	D	A	E	C	B	A	E	9
MC4	E	D	D	A	E	B	C	B	A	A	E	C	B			9
MC5	A	D	D	C	C	C	A	A	E	A	B	D	D	B	E	4
MP1	E	B	D	C	B	D	C	A	C	A	E	A	D	B	E	6
MP2	E	D	A	E	D	D	B	B	E	A	C	C	B	C	D	4
MP3	A	D	D	A	E	C	C	B	D	A	E	D	E	B	B	8
MP4	A	D	A	C	D	C	B	B	A	C	B	C	E	C	D	4
MP5	E	D	A	A	B	C	A	C	E	E	D	A	E	A	D	5
Key	E	D	D	A	E	C	C	B	C	B	E	C	A	A	E	15

Two of the five paper and pencil examinees did not write in the test booklet at all. Those examinees indicated that they felt they could do the problems in their head, and didn't need to use scratch paper. The other three paper and pencil examinees wrote in their booklet on 4, 7, and 10 problems. The five computer examinees wrote on their scratch paper on 1, 2, 2, 3, and 5 problems. Table 6 summarizes by item the number of paper and pencil and computer examinees showing work, and for problems with figures, the number marking on the figure (paper only) and the number redrawing the figure (computer only).

Computer examinees might also have been more inclined to use their calculator to store intermediate steps, rather than using scratch paper. For several items, we asked examinees about their calculator use in solving the items, but many examinees showed difficulty in recreating their calculator use on specific items, although they seemed able to recreate the process they went through to answer the item. These examinees might have been resolving the problem as they saw it the second time rather than recreating what they had done to solve it the first time.

Computer examinees that did not redraw figures on their scratch paper indicated that they did not redraw because the items were easy or because they had the picture in their head. One computer examinee confessed that "I think I should draw more often…I just don't draw. I never think to draw it again." Where computer examinees did redraw figures, it was to help them

visualize more in solving the problem. For Item 7, where it might have been helpful to draw a figure, a couple of paper and pencil examinees indicated that they didn't think about drawing a figure, although one examinee admitted in retrospect that drawing a figure might have helped. One computer examinee indicated that he drew a figure because "I plugged it in and I calculated that they'd want a graph from me," whereas another computer examinee said "I'm no drawer," in explaining why she didn't draw a figure for Item 7.

Table 6. Summary of Scratch Work for Math Items

Item	Item Type	Figure	Wrote on Booklet / Scratch Paper	Wrote on Figure / Redrew Figure	Drew Figure
1	PA	No	1P, 0C	N/A	N/A
2	PG	Yes	2P, 1C	1P, 0C	N/A
3	PG	Yes	1P, 0C	1P, 0C	N/A
4	IA	No	2P, 0C	N/A	N/A
5	EA	No	2P, 2C	N/A	N/A
6	TG	Yes	2P, 2C	2P, 2C	N/A
7	CG	No	3P, 2C	N/A	0P, 2C
8	IA	No	2P, 1C*	N/A	N/A
9	IA	No	1P, 0C	N/A	N/A
10	IA	No	1P, 1C	N/A	N/A
11	TG	Yes	0P, 0C	0P, 0C	N/A
12	PA	No	0P, 0C	N/A	N/A
13	PG	Yes	1P, 1C*	1P, 0C	N/A
14	CG	Yes	1P, 1C	1P, 1C	N/A
15	EA	No	2P, 1C	N/A	N/A

PA = Pre-Algebra
EA = Elementary Algebra
IA = Intermediate Algebra
PG = Plane Geometry
CG = Coordinate Geometry
TG = Trigonometry
P = Paper
C = Computer

*By accident, examinee MC1 was not given a calculator. The scratch work was multiplication only, which she indicated she would have done on her calculator if she had one.

Paper and pencil examinees were asked whether they were comfortable switching between writing on the test booklet and using their calculator. All said yes. Computer examinees were asked whether they were comfortable switching between writing on the scratch paper, using the mouse, and using their calculator. Again, all said yes. These questions did not get at whether the computer examinees felt they used their scratch paper to a different degree than if they had taken the test traditionally via paper and pencil. Computer examinees might have been unaware of a discrepancy between the scratch work they did do and the scratch work

they would have done in the test booklets had they taken the test by paper and pencil. One computer examinee stated, "I didn't write much, but it didn't seem like I needed to do so."

Reading

Based on previous experience, we anticipated that the booklet presentation of the Reading test might favor paper and pencil examinees overall, so that paper and pencil examinees might perform better on average on the test than computer examinees. Because the passage could never be seen in its entirety in the computer presentation, we anticipated computer examinees might have more difficulty both navigating throughout the passage (because of scrolling) and finding the information needed to answer than paper and pencil examinees.

Items requiring global understanding of the passage might be particularly difficult for computer examinees, if they search for a specific answer in the passage. Undirected searching (i.e., no line reference given) might require a lot of scrolling on the part of computer examinees. It might take more time to read a passage if scrolling is required in addition to just reading the passage. Booklet examinees do not have the extra navigational factor of scrolling added to the reading task. Items without line references that require specific knowledge might also be more difficult for computer examinees if the examinee has to review the passage at all to answer. Computer examinees might be less likely to exhibit positional memory then paper and pencil examinees (i.e., they might have a poorer memory of the layout of the passage and contents of sections of the passage), and might take more time to find relevant information in the passage.

Items containing line references might show a slight advantage toward computer examinees if they allow greater focus on the relevant portions of the passage, without the distraction of the noise from the rest of the passage. There might be a focus effect in general that could advantage computer examinees, if examinees are better able to focus on the limited information presented in the passage window. This advantage might be offset, however, by any scrolling required to get to the line(s) on computer. An item (with or without line reference) that refers to the same part of the passage as the previous item might be advantageous to computer examinees because the needed information is right there on screen.

The content of referenced lines might potentially be a factor in mode differences, if the line breaks differ across modes. Examinees might make different inferences if the content of the referenced line(s) differs even slightly across modes. Because of the structure of the computer presentation, fixing line breaks to be identical across the computer and booklet presentation in

this study would have resulted in longer passages and more scrolling on the computer, which could have been potentially disadvantageous to computer examinees. There might always be some degree of trade-off required between maintaining as close a representation of booklet material on the computer and the logistics of maintaining that representation. Page breaks might also be a factor, because they occur in the booklet, but not the computer presentation. If booklet examinees do not bother to read back to a prior page, this might be disadvantageous to paper and pencil examinees.

We asked examinees about their experience in answering 10 of the 20 Reading items. The examinees' interactions with the computer and booklet interfaces will be discussed for issues of line breaks, line references, and scrolling, along with issues related to items requiring global and specific levels of information. Table 7 presents a summary of examinee performance across items on the Reading test, by examinee ID. The shaded items are items that will be discussed relative to a test-specific issue.

Table 7. Reading Item Responses and Key

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
RC1	A	D	D	C	A	C	A	B	C	B	A	B	A	C	D	D	D	C	D	C	16
RC2	A	A	D	C	D	C	A	D	C	D	A	B	A	D	D	D	D	B	B	A	15
RC3	A	D	D	C	B	C	A	B	A	B	A	D	B	A	B	B	B	B	B	C	11
RC4	A	D	D	C	A	C	A	D	D	B	A	B	A	D	D	D	A	B	B	C	17
RC5	A	B	A	A	A	B	C	A	C	A	A	B	A	D	B	D	A	A	A	A	7
RP1	A	D	E	C	B	C	D	A	D	A	A	B	A	B	D	D	D	B	D	B	12
RP2	A	D	D	C	A	C	A	D	C	B	A	B	A	B	D	D	D	B	B	C	20
RP3	A	D	D	D	A	C	A	D	D	A	A	B	A	D	A	D	A	A	B	C	13
RP4	A	A	D	C	B	B	D	D	B	B	A	B	A	B	B	C	B	C	A	A	9
Key	A	D	D	C	A	C	A	D	C	B	A	B	A	B	D	D	D	B	B	C	20

Line Breaks. The stimulus in Item 14 contained a reference to a single line in the passage. The content of that individual line differed slightly across computer and booklet presentation. The stimulus was asking about the meaning of the term "blue". In the computer presentation, both the term "blue" and "blues" occurred on the referenced line, whereas in the booklet, only "blue" occurred on the referenced line. The question appeared difficult in general, because the terms "blue" and "blues" were both used in the surrounding sentences. Examinees that did not read the stimulus carefully might have been led to respond either way. Examinee RP1 demonstrated this potential source of confusion in her statement that "I made sure I reread that line twice – or the two lines twice – to make sure that I'm picking the blue one that they want." It is likely that many examinees would read more than just the sentence referenced to in

BEST COPY AVAILABLE

the stimulus. More careless examinees that do not read beyond the referenced line might not make that informed choice, but rather, go with what catches their eye. Having slightly different content on the referenced line across presentation modes could cause examinees within each mode to approach the item differently.

Table 8 summarizes results for Item 14. All of the paper and pencil examinees thought that the question was asking about "blue." Three of the four answered the item correctly (the fourth examinee chose the correct answer initially, but then changed her answer after finishing the test because she felt she had a better idea of the answer after reading more of the passage). Two of the five computer examinees thought the question was asking about "blue," two thought it was about "blues," and one was uncertain whether it was about "blue" or "blues." The two that thought it was about "blue" explained their answers as if the question was about "blues" rather than "blue." None of the computer examinees answered the item correctly. Allowing lines to break differently across administration modes, even if the content of the referenced line(s) differs only slightly, could have an unintended effect on how examinees respond to the item.

Table 8. Results for Reading Item with Line Break Issue

	Computer	Paper
Thought question was about "blue"	2	4
Thought question was about "blues"	2	0
Uncertain whether "blue" or "blues"	1	0
Answered Item correctly	0	3*
Number Taking Item	5	4

*Examinee RP3 initially answered item correctly but changed her answer

Item and Passage Relation. Item 16 presented an example of an item without a line reference that referred to information in the same part of the passage as the preceding item. We anticipated that such a relation might be advantageous to computer examinees because presumably the relevant section of the passage would already be in the window from the previous item. Table 9 shows the results for Item 16. Four of the five computer examinees recognized that the item referred to the same location as the previous item. One of those examinees was prompted by a key name he saw in the passage window; the others remembered from the previous item. One computer examinee (who answered the item incorrectly) thought she remembered the item being discussed in a different location, and scrolled to where she remembered it being located in the passage. Two of the paper and pencil examinees recognized

that the item referred to the same location as the previous item. One of those two did answer the item incorrectly, but expressed that he thought the item was one of the easiest on the test because of the location issue. Two of the paper and pencil examinees skimmed through the passage to find the key names associated with the item. They both answered the item correctly, but one admitted that the item was hard because she had to review about half of the passage and that it was time consuming to do so.

Table 9. Results for Item with Item and Passage Relation Issue

	Computer	Paper
Knew it referred to same Location as previous item	4	2
Had to search passage to find	1	2
Answered item correctly	4	3
Number Taking Item	5	4

Approach to Passage. Examinees typically approached the Reading passages in one of two ways. They either read a passage in its entirety before starting to answer questions, or they started answering the questions right away without reading an entire passage. Table 10 summarizes the approach to Passage 1 and items for the Reading test. Both computer and paper and pencil examinees showed tendencies to approach the passage in either way. Examinees that read the entire passage first might be more likely to answer from memory, or to remember specific locations in the passage to refer to for answers. Examinee RP2 discussed her strategy of reading the passage entirely and connecting every paragraph with a main idea. Examinees that did not read the passage first typically showed some trouble in general in finding both specific and global information within the passage on some items. Two of the paper and pencil examinees in particular, RP3 and RP4 discussed their confusion on several items because they had not read the passage completely. Computer examinees that do not read the passage first might be affected even further by the navigation required to move about and see the entire passage.

Table 10. Approach to Passage 1 and Items for Reading

	Computer	Paper
Read entire passage first	3	2
Didn't read entire passage before starting questions	2	2
Number Taking Passage 1	5	4

Positional Memory. Related to the approach to the passage is the issue of positional memory, namely, the ability to remember and place certain pieces of information with their location in the passage. When items referred to that information, examinees were able locate the placement of the information in the passage. Most of the Reading and Science examinees indicated that they experienced positional memory to some degree in both the computer and paper and pencil presentations. There was some question as to whether computer examinees would experience positional memory to the extent that paper and pencil examinees would, because the passage was not divided into any tangible units, such as pages, in the passage window. None of our data suggested that positional memory occurred to a lesser degree for computer examinees than for paper and pencil examinees. But this is something we continue to look for in our research. Although we have no evidence to support this, it might be that positional memory is functional for computer examinees only at the beginning and ending of the test. Positional memory might be more difficult for the middle of the test when there is no definite grouping associated with the material such as a page, or the beginning or end of the passage.

Science

The issues relevant to Science were very similar to the Reading issues, because the two tests shared a similar structure of passages with scrolling. We anticipated that computer examinees might perform more poorly than paper and pencil examinees on Science because of the scrolling and navigational issues associated with the computer interface. The Science item responses and key are presented in Table 11. Specific to Science was the capability of enlarging and moving graphics and tables. Three of the four computer examinees expressed frustration because they had to compare tables and figures that could not be viewed in the passage window at the same time. The enlarging capability would have allowed them to do so (it was possible to move enlarged figures anywhere on screen), but examinees did not enlarge because they thought they could see the figures fine. The computer examinees did not recognize that the enlarging capability would allow them move a figure and line it up with another figure.

Table 11. Science Item Responses and Key

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Total
SC1	D	B	A	B	C	C	A	C	B	D										10
SC2	D	B	A	B	C	C	A	C	B	D	A	C	A	B	A	D	B	D	C	14
SC3	D	B	A	D	D	C	A	C	B	D	C	C	A	B	B	B	A	B	B	12
SC4	D	B	A	B	C	B	A	C	B	D	D	D	A	B	B	D	A	D	A	18
SP1	B	C	C	B	C	B	A	B	B	D	D	D	A	B	B	D	A	D	B	13
SP2	B	C	D	B	C	C	B	C	B	D	C	A	B	B	B	D	A	D	C	11
SP3	D	B	A	B	C	C	A	C	B	D	D	D	A	B	B	D	A	D		18
SP4	C	B	D	D	C	A	A	C	B	D	D	D	A	A	B	D	A	B	B	12
SP5	D	C	A	C	C	C	A	C	B	D	D	D	A	B	B	D	A	C	A	16
Key	D	B	A	B	C	C	A	C	B	D	D	D	A	B	B	D	A	D	A	19

Global Test-Level Issues

In addition to asking questions about test-specific issues, we asked a number of questions pertaining to global issues that were relevant across all of the tests. The questions were designed to address both the examinees' behavior and their attitudes toward features of the computer and booklet presentations. Examinees were asked about their order of answering items, their attitude toward item review, item preview, and omits, their perceptions on scrolling, and their willingness to take a high-stakes test such as the ACT Assessment on the computer.

Order of Answering Questions, Item Review, and Item Preview

Slightly different rules were imposed across the computer and booklet presentations regarding order of answering items, item review, item preview, and omits. In the paper and pencil presentations, the examinees were allowed to move freely throughout all items in a test, answering in any order they chose, and were allowed to omit items if they chose (although all examinees were encouraged to guess on items that they did not know the answer). In the computer presentations, the examinees were required to answer all items and they were not allowed to review or preview items. For Math, the computer examinees were required to answer an item before moving on to the next item. For the passage-based tests, the examinees were allowed to move freely between items within a passage when answering, but were required to answer all items in the current passage before moving on to the next passage. They could not review a passage once moving on, or preview other passages.

Table 12 summarizes the order of answering items for the examinees. For the paper and pencil tests, examinees were asked about their behavior over all items. For the computer administered passage-based tests, the examinees were asked about their behavior on items within

passages. Most of the passage-based examinees indicated that they answered the items in order, and that they traditionally use that strategy on tests. Several noted that they did so because they were less likely to make a mistake filling in the answer sheet, or because it could be confusing to move between passages. Of the paper and pencil passage-based examinees that skipped around, Examinee EP3 skipped one item at the end of the test because he was running out of time, and went back to it after finishing the remaining questions. Examinee SP5 indicated that he usually would answer in order, but that he was trying a new strategy suggested by preparatory materials. For the computer examinees, examinee RC1 answered Item 10 of the second passage before Item 9. Examinee RC4 essentially answered in order, but went back to Item 6 in the first passage after answering Item 7 to change her answer, and went back to Item 9 in the first passage after answering Item 10 to change her answer. Examinee RP3 answered in order, but went back to redo a couple of items she was unsure of.

The Math paper and pencil examinees also demonstrated a propensity to skip around. Two of the Math examinees skipped one item and went back to it and answered it later. The other examinee indicated a strategy of skipping items that bother him, and going back to them later. One Math computer examinee indicated that although he thought it was easier to answer the items in order, he would definitely skip around sometimes in a paper and pencil test.

Table 12. Order of Answering Items

	In Order		Skipped Around	
Test	Computer	Paper	Computer	Paper
English	5	2	0	1
Math	N/A	2	N/A	3
Reading	3	4	2	0
Science	4	4	0	1

The ability to skip around and the ability to go back and review answers were very important concerns for the examinees participating in the study. Table 13 summarizes responses to the question of whether there was a point in a later passage (for passage-based tests) / item (for Math) where the examinee wanted to or did go back (for paper and pencil) to a previous passage/item. For the passage-based tests, the computer examinees uniformly indicated they did not want to go back to a previous passage once they had moved on to another passage. Only Examinee EC3 indicated a desire to do so, stating "sometimes I thought maybe I missed a question or something." The two paper and pencil Reading examinees that indicated they went

BEST COPY AVAILABLE

back to a previous passage did so to review answers after completing the test. For the Math test, three examinees each in both the paper and pencil and computer modes expressed a desire to go back, while two each in both modes said they did not want to go back.

Table 13. Did Examinee Want To Go Back To Previous Passage/Item After Moving On

	Computer		Paper	
Test	Yes	No	Yes	No
English	1	4	0	3
Math	3	2	3	2
Reading	0	5	2	2
Science	0	4	0	5

The examinees were asked further about their feelings about not being able to go back to previous passages/items (computer) or how they would feel if they were not able to go back to previous passages/items (paper). Their replies are summarized in Table 14. Many of the computer passage-based examinees indicated that they did not mind not being able to go back because the passages were not related to each other. Some viewed the passages as separate little tests and were comfortable from previous experiences about not being able to go back to an earlier test. Others indicated that they would only want to go back if something in the current passage gave them a clue to an answer in an earlier passage. Two Science paper and pencil examinees indicated that they thought they would take more time on questions because they could not go back, and thought that would hurt their performance. One Science paper and pencil examinee did not know how to respond because he was trying a new strategy for the first time.

Table 14. Feelings About Not Being Able To Go Back To Previous Items

	Computer		Paper	
Test	Dislike	No Difference	Dislike	No Difference
English	1	4	1	2
Math	2	3	3	2
Reading	0	5	3	1
Science	2	2	3	1

In general, the feeling on the passage-based tests was that it was not that bothersome for examinees to be unable to go back to a previous passage. What examinees did seem to want, however, was to be able to review their answers after completing the test. Many of the "No

Difference" responses were conditional on the fact that they be able to review their answers at the end of the test. The paper and pencil passage-based examinees seemed to feel a little more strongly than the computer examinees about not being able to go back. Had they taken the test in the computer mode, they might have been more inclined to be indifferent about not being able to go back.

The computer examinees were also asked if they would have checked over their answers at the end of the test if they had been able to. With the exception of examinee EC2 who indicated he had already checked each passage carefully before moving on, they all indicated they would have checked their answers. Comparing this to what the computer examinees actually did showed some discrepancies. Again, with the exception of EC2, none of the computer passage-based examinees actually checked any of their answers in Passage 2 after completing the passage and before ending the test (even if they had time remaining). So the examinees might think they will do one thing in a given situation, but in reality will do something else when they are actually in that situation.

The Math examinees expressed more dissatisfaction about not being able to go back to previous items, both for computer and paper and pencil examinees. In general, the Math examinees wanted freedom to skip items and go back, and to be able to review their answers at the end of the test. Based on examinee comments, it seems a fairly common practice for Math paper and pencil examinees to skip difficult items and go back to them later in the test, so that examinees would generally be bothered by not having that freedom in the computer administration.

Item preview was less of an issue for the examinees. They generally expressed an interest in being able to preview items on tests only to see how many items remain and the level of difficulty of remaining items, in order to help them gauge how much time to spend on the current item/passage. Overall, it might have been more difficult to allocate time on the computer administration than the paper and pencil administration. Problems in gauging time might be eliminated to some degree if the examinees were continually aware of the number of items/passages remaining. Other examinees might continue to have difficulty gauging their time without seeing the difficulty of the remaining items. The examinees did not seem to mind too much being forced to give an answer rather than omit on the computer for the passage-based

tests, because the instructions tell them to guess when they don't know the answer. On the Math test, again the examinees wanted the capability of omitting an item in order to return to it later.

Scrolling

There was some concern on our part that the act of navigating through the computer environment, particularly scrolling to read passages, might have required more time to complete the test on computer than on paper and pencil. The item responses in Tables 2,5, 7, and 11 show the not completed rate for the English, Math, Reading, and Science tests, respectively. The paper and pencil examinees all completed the test, with the exception of examinee SP3, who did not complete the last item. On the computer, examinee EC1did not complete the last seven English items, examinee MC4 did not complete the last two Math items, while examinee SC1 did not complete the last nine Science items. All Reading examinees completed the test in the computer mode. We asked the computer examinees taking the passage-based tests to compare the test to the same test administered conventionally, and to determine whether scrolling on the computer test helped or hurt their performance, or had no effect. There was no scrolling capability in the Math computer administration. Table 15 shows their responses. For the computer examinees that did not complete the test, examinee EC1 thought scrolling had no effect and felt having to scroll to read was made up for by the quicker answering speed. Examinee SC1 thought that scrolling hurt her because she did not get as much done. She acknowledged that on paper she would keep her finger on the spot when comparing two things, and thought that would make it easier to compare than on the computer.

Two English examinees said scrolling helped; one because she thought the automatic scrolling in the passage window helped give more focus, the other thought the quicker answering speed on the computer made up for the effect of scrolling. For the examinees that thought they were hurt by scrolling, they indicated that not having the entire passage in view was problematic. Examinee RC5 said that it was easier on paper to get some order to the paragraphs and remember where things were. The other examinees indicated that they liked the focus the passage window gave them, and the presentation of one item at a time. Examinee RC4 liked the computer presentation because it was not as overwhelming as the booklet presentation of the passage and items.

Table 15. Effect of Scrolling on Computer Performance

Test	Helped	Hurt	No Effect
English	2	1	2
Math	N/A	N/A	N/A
Reading	0	1	4
Science	0	2	1

Willingness To Take High-Stakes Test On Computer

To complete the interview, examinees were asked about their willingness to take a high-stakes test such as the ACT Assessment on computer. Table 16 summarizes their preferences. The examinees that said "yes" in general liked the ease of taking the test on computer rather than paper and pencil. They liked not having to use pencils and not having to bubble in answers. The examinees that said yes conditionally ("Yes If...") uniformly indicated that they would take a computerized version if they had the same freedom as the paper and pencil test to go back and see previous items and answers. The examinees that said "no" preferred the paper and pencil in general because they were more comfortable and more familiar with that style of testing. Examinee RC5 summed up the consensus feeling with her statement that "I would do it paper and pencil because that's what I'm used to with taking tests." The perceived lack of control might be difficult to overcome when tests are first administered via computer, because examinees will have the expectations that they have learned from years of testing via paper and pencil. Examinees that took the test on computer in our study might have been more inclined not to rule out taking the test via computer than examinees that took the test on paper. Examinees might require a certain level of experience with testing on computer before they are comfortable with it.

Table 16. Willingness To Take High-Stakes Test On Computer

	Computer				Paper			
Test	Yes	Yes If...	No	No Preference	Yes	Yes If...	No	No Preference
English	2	1	1	1	1	2	0	0
Math	1	4	0	0	2	1	2	0
Reading	3	0	1	1	0	1	4	0
Science	1	0	3	0	2	0	3	0

Discussion

There were a number of factors in the study that could have affected the observed outcomes. First, the examinees that participated in the study appeared to be fairly computer literate. Overall, it is difficult to say whether these examinees were more computer literate than the general population, and if so, whether less computer literate examinees would have interacted differently with the interface. But responses to the interview questions indicate that examinees that are not comfortable with computers are probably unlikely to take a high-stakes test via computer as long as it is offered by paper and pencil.

A second factor was examinee motivation. Because there were no stakes attached to the test scores, some examinees might not have been very motivated, or might not have approached the test exactly as they would have under actual conditions (i.e., in terms of reviewing and how much time they devoted to individual questions). Several of the examinees did state they were nervous about the time, or felt rushed because they were running out of time, which seems to indicate that they took the test fairly seriously. Further, the attitude of the study participants was very good. Many examinees seemed to enjoy the experience and expressed an appreciation of being given an opportunity to discuss their opinions and impressions about testing and the particular test they took.

A third factor was the use of a follow-up interview, where students were not asked about individual items until after they had completed the test. With after-the-fact questioning, it was difficult to know whether the examinees told us what really happened, or whether they answered the items anew as they reviewed them. Our findings were limited both by how well the examinees were able to remember the process they went through, and by their ability to describe that process to us. Several additional examinees tested under untimed conditions using a think-aloud format, where they were directed to think aloud what they were thinking and doing while taking the test. We did not report the think-aloud results because we felt that it was important for the examinees to take the test under timed conditions. Taking a test in an untimed situation could cause examinees to react and interact differently with the presentation features we wanted to assess.

One shortcoming of the follow-up interview approach was that examinees exhibited at times signs of uncertainty as to which answer they had chosen and why, and what they did while taking the item. Further, the examinees seemed more able to remember what they did to answer

an item mentally than what they did physically in navigating throughout the computer environment to answer items. For example, computer examinees often talked about scrolling for an item when they did not scroll at all while responding to that particular item. For the computer examinees, we were able to verify the accuracy of their memory because we had videotape coverage, but were unable to do so for paper examinees except by speculation.

The time in which the study took place was also somewhat problematic, occurring at the end of summer vacation. The Math examinees all did fairly poorly, and many stated that they had either forgotten formulas or how to solve problems over the summer. Many of them thought they would have performed better had the study taken place during the school year. Examinees from the other content areas might also have been somewhat out of practice at taking tests due to the long break.

The short testing time might also have been a limitation of the study. We are interested in the effects of fatigue on computer examinees, and whether examinees get more tired testing on the computer than by paper and pencil because of the visual strain. The short testing time in the study did not really allow us to get at that issue. We purposely chose the short testing time to ensure that examinees would be able to remember the test items and what they had done to answer them. A number of examinees did report feeling rushed for time. So although there might not have been fatigue, there might have been some element of speededness in taking the test. We took care to set the test time, though, so that speededness would not be an issue all throughout the test.

A last influential factor could have been the characteristics of the individual interviewers. At times, we noted instances of potentially leading behavior from the interviewers. Sometimes an interviewer used wording and terminology that differed from the prepared script, that might have indicated to the examinee what we were hypothesizing. In those cases, the examinees generally did not appear to be led. We also observed that sometimes an interviewer answered for the examinee what they thought the examinee was trying to say, rather than waiting for the examinee to say it, or prompting the examinee for a response. This interviewer behavior could have resulted in the examinee agreeing to a statement that they might not have made, had they spoken freely for themselves.

Because the sample sizes were very small, we cannot use the results of the study to identify any trends, but rather only as an indication of how an individual examinee might react to

the mode presentation features. The observations we have made for the study sample do, however, highlight issues that test developers might want to consider in determining how to present a test, in either a computer or paper and pencil administration mode. We have identified several factors that might lead an examinee to respond to more than just the item content when giving their answer, such as page and line breaks, passage and item layout features, highlighting, and item characteristics. Other factors include navigational features such as scrolling, item review, item preview, and omit capability. The effects of these factors on performance might be dependent upon the administration mode and the features of that administration mode. Some factors might be more controllable than others. Test developers should be conscious of these factors when making formatting decisions, particularly to minimize mode differences where dual-platform testing is employed.

Examinee characteristics contribute to many of the observed mode effects. For example, careless examinees are much more likely to be led astray by presentation features than careful examinees. But there might be some presentation features that lead even careful examinees to be tripped up. It is our task to remove those factors, where possible, that could lead careful examinees to be misled. The use of different line breaks leading to the "blue" vs. "blues" confusion in the Reading test is a primary example. We should not, however, be held responsible for examinee carelessness (i.e., examinees that do not read an item stimulus in English when they should, or examinees that do not read an entire underlined portion in English because of a page or line break). It is the test developer's responsibility to minimize presentation differences wherever possible, but ultimately, the examinee must be responsible for following the test directions.

Unfortunately, a timed test might be a primary cause of examinee carelessness. When tests are timed, some examinees might use timesaving devices while taking the test. One timesaving device we observed was that examinees skipped reading things they deemed unnecessary. For example, in the English test, an examinee might not read the entire sentence containing the underlined portion, or the sentences surrounding it, if he or she decides it is not necessary. Or, if a stimulus exists, the examinee might attempt to answer without reading the stimulus, or without reading the stimulus in its entirety. Not reading fully can lead to trouble if the examinee does not get the full gist of what the required task is. Some study participants purposely chose not to read a stimulus where it existed, because of the timed nature of the test.

Perhaps this behavior was not carelessness on the part of the examinee, but it had the same effect as carelessness, in that the examinee might have missed important or necessary information. Removing the time factor might help eliminate some of the hypothesized sources of mode differences such as page breaks, highlighting, or scrolling, but might not affect others such as line breaks, or passage layout features. And some examinees might be careless regardless of the timing of the test.

Every examinee has to be viewed as a unique entity whose approach to a test is affected by his or her experiences, characteristics, and expectations. Because of each examinee's uniqueness, it is impossible to predict how an individual examinee will react to an item and the presentation features associated with it. To some degree, examinee factors can be controlled through educational materials about the test and the administration features. Initially, examinees will expect what they have known in the past. As they gain experience within a new administration mode, those expectations will change. Of greater concern, are the examinee factors that cannot be controlled through educational materials. As we move further into the realm of computerized testing, test developers need to be cognizant of item characteristics and the effect that formatting and presentation choices could have on an examinee's response. If computer presentation features are so dominant that the examinee is inclined to react in a different manner than had the item been presented in a paper and pencil administration, that is a problem. All care should be taken in the test development process so that we can be confident that an examinee is responding to item content only, and not to inherent features associated with presenting the item in an administration mode.

References

Davey, T., & Nering, M. (1998, April). *Evaluating and insuring measurement precision in adaptive testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Davey, T., Nering, M., & Thompson, T.D. (1997, June). *Realistic Simulation Procedures for Item Response Data*. Paper presented at the Psychometric Society meeting, Gatlinburg, TN.

Fan, M., Thompson, T.D., & Davey, T. (1999, April). *Constructing adaptive tests to parallel conventional programs*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.

Hsu, Y., Thompson, T. D., & Chen, W. (1998, April). *CAT item calibration*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, April.

Mazzeo, J., Druesne, B., Raffeld, P.C., Checketts, K.T., and Muhlstein, A. (1991). *Comparability of computer and paper-and-pencil scores for two CLEP general examinations* (Research Report No. 92-14). Princeton, NJ: Educational Testing Service.

Miller, T., & Davey, T. (1999, April). *Principles for administering adaptive tests*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.

Nering, M., Thompson, T. D., & Davey, T. (1997, June). *Simulation of realistic ability parameters*. Paper presented at the Psychometric Society meeting, Gatlinburg, TN.

Neuman, G., & Baydoun, R. (1998). Computerization of paper-and-pencil tests: When are they equivalent? *Applied Psychological Measurement, 22*, 71-83.

Parshall, C. G., Davey, T., & Nering, M.L. (April, 1998). *Test development exposure control for adaptive testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Parshall, C.G., & Komrey, J.D. (1993, April). Computer testing versus paper-and-pencil testing: An analysis of examinee characteristics associated with mode effect. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Reckase, M. D., Thompson, T. D., & Nering, M. (1997, June). *Identifying similar item content clusters on multiple test forms*. Paper presented at the Psychometric Society meeting, Gatlinburg, TN.

Schaeffer, G.A., Reese, C.M., Steffen, M., McKinley, R.L., & Mills, C.N. (1993). *Field test of a computer-based GRE general test* (Research Report No. 93-07). Princeton, NJ: Educational Testing Service.

Spray, J.A., Ackerman, T.A., Reckase, M.D., and Carlson, J.E. (1989). Effect of the medium of item presentation on examinee performance and item characteristics. *Journal of Educational Measurement*, *26*, 261-271.

Thompson, T. D., & Davey, T. (1999, April). *CAT procedures for passage-based tests.* Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.

Thompson, T. D., Davey, T., & Nering, M. (1998, April). *Constructing adaptive tests to parallel conventional programs.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Thompson, T. D., Nering, M., & Davey, T. (1997, June). *Multidimensional IRT scale linking without common items or common examinees.* Paper presented at the Psychometric Society meeting, Gatlinburg, TN.